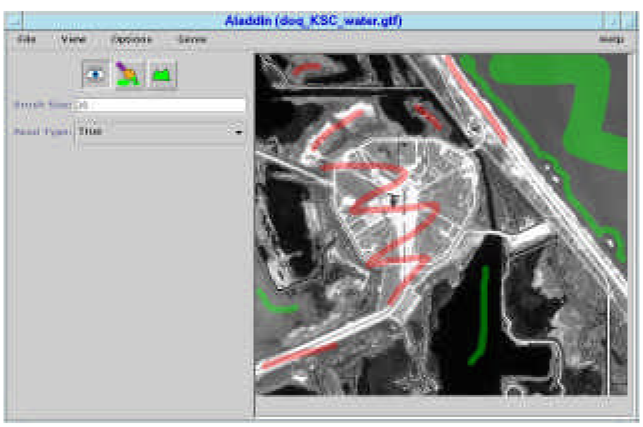
Rapid Feature Identification with Genetic Algorithms

Points of Contact: John J. Szymanski (505) 665-9371, szymanski@lanl.gov
Jeffrey J. Bloch (505) 665-2568, jbloch@lanl.gov

Los Alamos National Laboratory
Space and Remote Sensing Sciences

A user of remotely sensed imagery faces the problem of finding features of interest within the imagery in the face of ever-growing quantities of incoming raw data, coupled with an ever-shrinking staff of trained analysts who can interpret that data. Significant effort has gone into developing dedicated Automatic Feature Extraction (AFE) tools that can identify particular features of tactical and strategic interest in remotely-sensed image data. However, these tools often suffer from the problem of being too specialized - while a particular AFE tool may be good at classifying vegetation or identifying impassable rivers in a Central European setting, it may fail completely when faced with a similar problem in a Mediterranean littoral or Mid-East desert landscape. The Rapid Feature Identification Project (RFIP) uses a different strategy: rather than design AFE tools by hand, the RFIP exploits recent advances in machine learning methods to develop such tools semi-automatically, using only a small collection of hand-classified images as a starting point. This tool will make it possible for an analyst to produce an AFE tool suitable for a given day's data set, tuned for a specific feature.

GENIE (Brumby, et al 1999; Theiler et al 1999) is an evolutionary computation (EC) software system, using a genetic algorithm (GA) to assemble image-processing tools from a collection of low-level image operators (e.g., edge detectors, texture measures, spectral operations, various morphological filters). Genie is typically being applied to multispectral and hyperspectral data sets. Each candidate tool generates a number of intermediate feature planes, which are then combined using a supervised classifier (currently a Fisher discriminant and intelligent threshold function) to generate a final result mask. A population of candidate tools is generated, ranked according to a fitness metric measuring their performance on some user-provided training data, and fit members of the population permitted to reproduce. Several standard fitness metrics have been implemented, including Euclidean distance and Hamming distance. The process cycles until the population converges to a solution, or the user decides to accept the current best solution or to change the training data. The burden of low-level programming is thus shifted to the genetic algorithm, leaving the analyst free to concentrate on the critical task of making judgements. GENIE is free to ignore the spatial information in the image and rely wholly on spectral operations and the supervised classifier, but in practice GENIE will construct integrated spatio-spectral algorithms. These have been shown to be effective in looking for complex terrain features, such as golf courses (Harvey et al, 2000).



As with all machine learning systems, performance depends crucially on the provision of a sufficient quantity of training data, and supplying this data is typically a major challenge. For

GENIE, training data is provided via an intuitive and highly graphical interface, featuring "point-and-click" functionality. The user is able to influence the evolution of algorithms by providing additional information, and by interactively providing additional training data. GENIE can derive multiple features for the same scene to produce terrain classifications (Brumby et al 2000). GENIE has been applied to Landsat 7 ETM+, Landsat 5 TM, and MODIS Airborne Simulator multispectral imagery, AVIRIS hyperspectral imagery, and to standard color/infrared aerial photography. GENIE is currently being tested for use with high-resolution panchromatic and laser altimeter data from NASA planetary remote sensing imagery.

The system employs both spectral and spatial image analysis techniques in combination, and can in principal simultaneously exploit data from different sensors (e.g., optical imagery plus multi-spectral imagery plus altimeter data or digital elevation models). The ability to combine diverse datasets requires that the data be co-registered, which requires use of some other package (e.g., RSI's ENVI or ERDAS Imagine commercial software packages). The code is written in a combination of Perl (for the GA), Java (for the graphical user interface, or GUI), and IDL, and augmented by our own C/C++ libraries for fitness evaluation of candidate tools. The code was developed in an Intel/Linux environment.

The prototype system typically requires a few hours to evolve a high-fitness image-processing algorithm running on a single, fast Linux/Intel workstation. The GENIE system is parallelizeable and scalable, and we have been developing a version of our system for a cluster of 10's to 100's of commercial off-the-shelf workstations.

References:

"Investigation of Image Feature Extraction by a Genetic Algorithm", S.P. Brumby, J. Theiler, S.J. Perkins, N. Harvey, J.J. Szymanski, J.J. Bloch, and M. Mitchell, Proceedings of SPIE 3812, 1999, pp. 24--31.

"Evolving Retrieval Algorithms with a Genetic Programming Scheme", J. Theiler, N.R. Harvey, S.P. Brumby, J.J. Szymanski, S. Alferink, S. Perkins, R. Porter, and J.J. Bloch, Proceedings of SPIE 3753, 1999, pp.416--425.

"Finding Golf Courses: The Ultra High Tech Approach", N.R. Harvey, S. Perkins, S.P. Brumby, J. Theiler, R.B. Porter, A.C. Young, A.K. Varghese, J.J. Szymanski, and J.J. Bloch. To appear in the proceedings of EvoIASP 2000: The Second European Workshop on Evolutionary Computation in Image Analysis and Signal Processing, Edinburgh, April 17th, 2000.

"A genetic algorithm for combining new and existing image processing tools for multispectral imagery" S.P. Brumby, N.R. Harvey, S. Perkins, R.B. Porter, J.J. Szymanski, J. Theiler, and J.J. Bloch. To appear in the proceedings of Aerosense 2000: SPIE's 14th Annual International Symposium on Aerospace/Defense Sensing and Controls, Orlando, Florida, April 24-28, 2000.